COMPUTER SCIENCE 3753
Assignment #4

Points: 100 **Weight**: 3%
Due: Friday, October 19, 2018 at 11:55 pm in BlackBoard
Note: Late assignment will not be accepted without instructor's pre–approval.

Hand in a IPython notebook **yourNmae-homework04.ipynb**, with appropriate markdown cells for descriptions and comments. Write Python code in code cells to complete the following tasks. ***This homework must be completed individually.***

1. [10] Download the UCI Automobile Data Set from https://archive.ics.uci.edu/ml/datasets.html. You can either download manually or using your own python code.

2. [10] Load the automobile data into a Pandas DataFrame named df. Name the columns and convert column data types according to the description on the data set webpage.

3. [10] Create a copy dataFrame named df2 from df. In df and df2, replace missing values and unknown values by meaningful values: the mean for numerical columns and the mode for categorical columns.

4. [20] Add a new column loss-percent to df2. Let m and M be the minimum and maximum in column normalized-loses, respectively. For each row, set the loss-percent to $(nl - m)/(M - m)$ with 4-digit precision, where nl is the normalized-loses on that row.

5. [20] Add another new column binned-engine-size to df2. For each row, the binned-engine-size must be the name of the bin that contains the engine-size. You must define appropriate equal-width bins and create meaningful bin names. For example, you may set bin width to 10 and create enough number of bins cover the range between min and max values.

6. [20] Draw a horizontal bar chart for each of the four attributes of df2: normalized-losses, loss-percent, engine-size, and binned-engine-size.

7. [10] For each of the dataFrames df and df2, print a summary table that describe statistics for each column. One table for categorical columns and one table for numeric columns. Make sure the newly created columns of df2 are included.